Appears in the Proceedings of SPIE 1994, Stereoscopic Displays and Virtual Reality Systems. Vol. 2177, pp. 365-371.

Hands-off interaction with menus in virtual spaces

Rudy Darken

The George Washington University
Department of Electrical Engineering and Computer Science
and
The U.S. Naval Research Laboratory
Washington, D.C.
darken@enews.nrl.navy.mil

ABSTRACT

As dissimilar as virtual environment interfaces are to traditional desktop interfaces, one common element will always be the need for selection. Menus are often used for this type of interaction to minimize memorization requirements. However, the direct analogy of two-dimensional menus to three-dimensional applications has not been widely accepted due to complications concerning pointing tasks in virtual space.

In this paper, we report results of the use of a new technique for menu display and interaction which involves the display of menu items on a two-dimensional overlay onto the three-dimensional world. The proposed technique displays the textual menu items on a viewplane which moves relative to the user so that text will always directly face the eyes. The menu items are selected via a speech recognition system. Advantages of this technique include ease of readability and trivial interaction with the menu. The hands are always free to be used for other tasks. A number of usability issues are discussed including the placement of the menu bar, font type and size, menu availability, the use of highlighting, and the depth and breadth of the menu tree. These findings show that this metaphor is an effective technique for not only menus in virtual spaces but for many uses of text in the three-dimensional domain.

1. INTRODUCTION

One of the defining elements of virtual environments is their extensive use of direct manipulation interaction techniques. But for commands and controls which do not have a direct manipulation counterpart, a menu selection style of interaction may be required. In its simplest form, the menu presents a series of choices on the display from which the user makes a selection. This technique can minimize or eliminate training and memorization of complex command sequences [10].

The use of menus in textual and two-dimensional desktop style interfaces has been investigated extensively [1, 6, 8, 11, 12, 13]. But with the advent of virtual environment technology and immersive displays, we are now interested in new ways of providing this same functionality. What problems does this new interaction paradigm pose for menu display and interaction?

A human-computer interface designer must design interaction techniques that minimize the work required by three types of basic human processes: perception, cognition, and motor activity [4]. These fundamental requirements apply to immersive interfaces as well as desktop graphical user interfaces. Extending these requirements to a menu system, we are primarily interested in the following issues:

- Clear visibility
- Simple interaction techniques
- Efficient menu structure
- Account for both novice and expert use

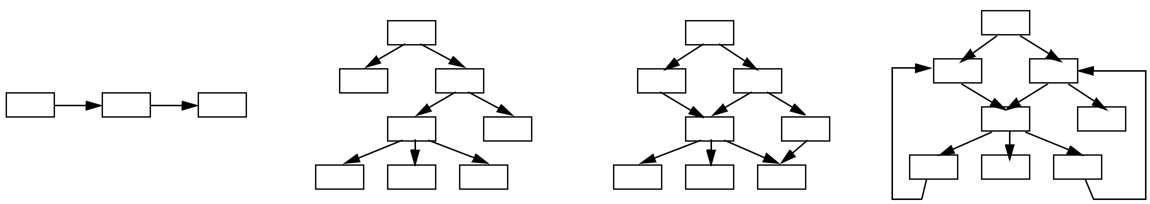
Our objective is to simplify menu selection, empowering novice users with the same tools available to experts. At the same time, the expert user should not be burdened with superfluous information. The menu selection technique described in this paper breaks the artificial bond interface designers have created between menu selection tasks and pointing tasks. We will examine the shortcomings of using a direct analogy from two-dimensional, pop-up style menus to three-dimensional spaces. A new technique is introduced which overcomes these difficulties by providing readable, efficient menu hierarchies from which selection tasks are trivial. By enforcing that execution of inherently two-dimensional tasks take place in two-dimensional space, the task can be simplified and made more efficient.

2. BACKGROUND

2.1. Menus in human-computer interfaces

Menus have been commonly used in product design for many years. We are all familiar with menus on items such as vending machines, video cassette recorders and microwave ovens. In their earliest use in computing systems, menus were used in Teletype machines and thus were required to be entirely textual. Later, with the capabilities offered by raster graphics and interaction devices for two-dimensional pointing such as the mouse, user interfaces became graphical rather than textual. Menu techniques evolved accordingly. The two predominant menu styles of today are pop-up and pull-down menus. Pop-up menus are activated on request at the cursor location while pull-down menus are activated from the menu bar containing the titles of the primary menu categories. The “look-and-feel” of these two techniques is very similar. They both typically list items vertically in a rectangular form and are dismissed after use.

Menu techniques in graphical user interfaces have been enormously successful primarily due to their suitability to the desktop metaphor. They fulfill all the requirements for effective use. They are clearly visible. The interaction requires only knowledge of the standard mouse interaction techniques (i.e. point, click, and drag). Keyboard accelerators have been developed to facilitate expert users. For most common menu selections, a keystroke allows the expert user to continue typing without reaching for the mouse. A number of configurations have been shown to be effective for use with menus. In particular, linear sequences, tree structures, and acyclic or cyclic networks [10] can be used to structure the menu hierarchy for optimized performance (See figure 1).



a. b. c. d.

Figure 1: Menu configurations can be a) linear, b) tree structures, or c) acyclic or d) cyclic networks.

2.2. The 2-D to 3-D analogy

With all the success graphical menus have had within the context of the desktop metaphor, their direct counterpart in the three-dimensional arena has not been as widely accepted. There are several reasons for this. Pointing techniques in three-dimensional spaces are much more complex and error prone than in two-dimensional spaces. Typically, a six degree of freedom magnetic tracking device is used in combination with an instrumented glove device to enable pointing to three-dimensional banks of textual items [5]. Magnetic tracking devices, even under the best of conditions, are noisy and require extensive filtering. This makes pointing difficult especially if the user is at some distance from the target or if the target is small.

Another issue is the use of stereoscopy for depth perception. Target acquisition in three-dimensional space requires the preservation of visual depth cues. Often, immersive displays do not adequately produce images from which correct depth information can be found. Without these cues, the task of pointing (intersecting a ray projected from the finger with a target) is exacerbated.

Hand gestures are used to interact with the menu in performing tasks such as activating the menu, selecting an item in the menu, or triggering the selected menu item. These gestures are not intuitive and require some level of training and practice for efficient use. There is also a visibility and readability problem since text is inherently two-dimensional but is represented in three dimensions. The user can move too far away for the text to be legible or might be at some angle from which the text is skewed making it unreadable.

Most importantly, the use of direct pointing for menu selection in virtual space is awkward due to its inefficient use of the hands. In the real world, the hands are used for grasping and working as well as for pointing. In virtual environments, the hands should be used primarily for direct manipulation-related tasks. For example, the hands should be used to point to some object to select it or to grasp virtual objects in the space for movement or alteration. Since the hands are a fundamental mechanism by which humans interact with their world, they must be utilized efficiently. Without the use of effective tactile feedback, grasping of virtual objects can be difficult. After grasping a virtual object for manipulation or inspection, it is awkward to have to put it down in order to make a menu selection.

The six fundamental types of interaction tasks described by Foley, Wallace, and Chan [4] apply to virtual environment interfaces as well as conventional graphical user interfaces.

- Select: Make a selection from among a number of alternatives
- Position: Specification of a spatial location
- Orient: Specification of a spatial orientation
- Path: Specification of position and orientation with respect to time
- Quantify: Specification of a value
- Text: Language-based textual input

Panel style menus in virtual environments require that a selection (often textual) be made from a bank of items using position and orientation of the hand. However, position and orientation tasks are prevalent in virtual environment applications more so than in two-dimensional graphical user interfaces. This is true because of the role of space in the virtual world. The menu technique described here attempts to decouple the task of menu selection from position and orientation subtasks thus allowing menu selection to occur in parallel with other primary tasks.

3. DISCUSSION

3.1. The hands-off menu selection technique

This technique for menu display and interaction involves the display of menu items in a two-dimensional overlay onto the three-dimensional world.The technique displays the textual menu items on the viewplane which moves relative to the viewpoint so that text will always face the eyes directly eliminating any readability problems. Figure 2 illustrates the principle concept behind the technique. The viewplane with two-dimensional text items is placed between the user and the three-dimensional virtual world. Figure 4 shows a typical view from the user's perspective. Menu hierarchies

Figure 2: The menu is drawn on a two-dimensional viewplane over a three-dimensional world of three objects.

are implemented as depths into the menu tree where leaf nodes are commands and nonleaf nodes are submenus. Initially, the main menu is visible. If a selection is made of an item which is not a leaf in the menu tree, the menu list is replaced by the new submenu. This continues until either a leaf is selected or the process is aborted. At any time, the user can return to the base menu or jump up one level in the menu hierarchy.

There are two key components to the technique:

1. The display must be high resolution for the text to be legible at a relatively small font size where the user can simultaneously view the three-dimensional world behind the text. Our implementation uses a BOOM2C by Fake Space Labs, Menlo Park, California which yields 1280 lines of horizontal resolution in its monochrome mode.
2. A speech recognition system is used for menu item selection. The speech recognition system must have isolated word capability. The recognition features can be simple but the speed and accuracy rate must be very high. This may seem to be a stumbling block for this technique but with the recent announcement by Silicon Graphics Inc. that in

the near future speech recognition functionality will be a standard component of their systems, it is clear that the use of voice in computer applications will become commonplace [2].

The technique is implemented as overlays on the graphics window. Each menu and submenu is defined and handled by the font manager for consistency. The operator of the system wears a lightweight microphone connected to the computer running the speech recognition software. This microphone can also be directly attached to the BOOM2C if it is undesirable for the operator to wear a device. The simulation runs separate from the speech recognizer. As spoken words are recognized, the speech recognition system sends an XEvent structure [7] across the network to the simulation process which handles it accordingly either performing an action or changing the overlay plane to a new submenu.

3.2. The use of text in high resolution immersive displays

The earliest LCD-based head-mounted displays sacrificed resolution for light weight. This made the use of text impractical if not impossible. One of the earliest high resolution immersive displays was the BOOM2 and later, the BOOM2C which offers a resolution of 1280x1024 pixels per eye in monochrome mode. With this display, the use of graphical text in the interface is possible. However, due to the optics involved, the effective resolution (i.e. the area of the display in which text is visible) is considerably less than the full resolution.

The visual field is most clear in the center of the display gradually degrading to black in a radial elliptic fashion. The effective resolution of the BOOM2C is illustrated in figure 3. The absolute center of the display is pixel location 640,

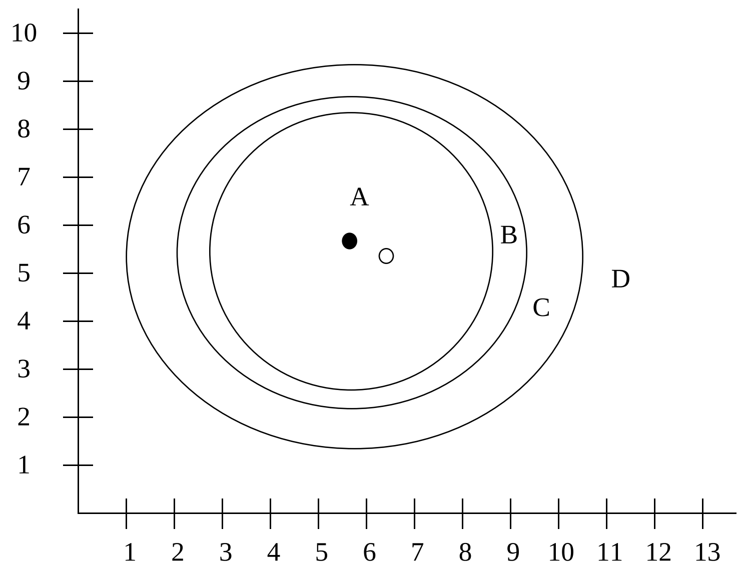


Figure 3: The effective resolution of the BOOM2C in hundreds of pixels.

512 (marked with a white circle). The optics alter this location to approximately 578, 565 (marked with a black circle). The three quasi-ellipses in the figure are centered at this location. Region A is the best readable area. Text which was found to be readable in the center of the display was readable throughout this region. In region B, text begins to blur. If used here, the font size must be enlarged. In region C, graphical elements are still visible but detail gradually vanishes to the edge of region D where the display is black. A typical display is shown in figure 4. The elliptic shading is simulated here to show the effects of the optics on the resolution of the display.

A number of font types and sizes were also tested in the display for readability. We found that typically, a minimum font size of 18 points is acceptable with 24 points optimal if display space permits. Helvetica is most readable but any standard font family such as Times-Roman, Bookman, or Palatino is satisfactory. Bold type did not noticeably improve readability.

4. RESULTS

The advantages of this technique are numerous. There is no readability problem if the text is placed in the view properly under the constraints placed by the optics. The effective resolution must be considered so that the text is placed where it will not only be in view but will also be clear. The effective resolution may differ between display devices and optic lenses.

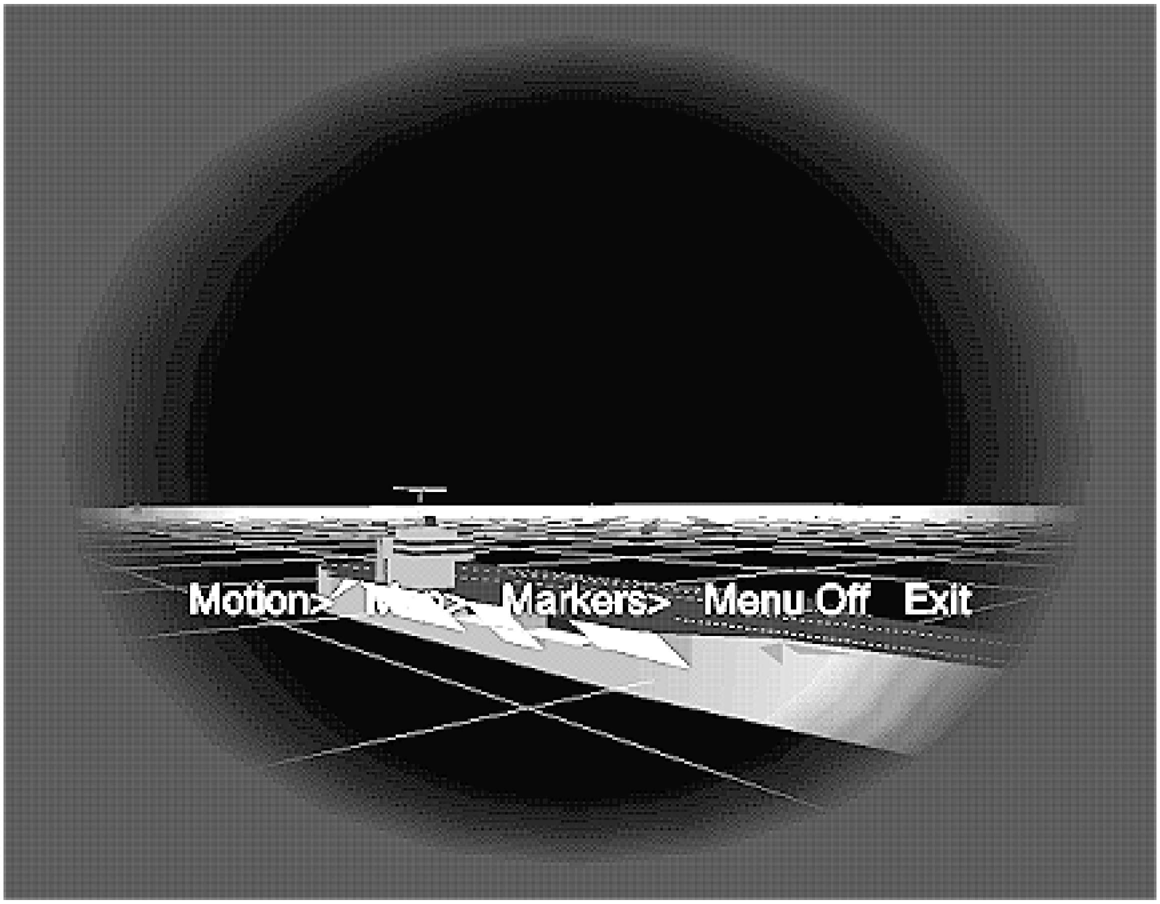


Figure 4: A typical display using the hands-off menu selection technique.

The interaction with the menu is trivial. The word displayed is the word to be spoken to trigger the command or submenu. The hands are always free to be used for other tasks such as grasping or pointing. This also allows menu item selection to be used in compound dialog structures. For example, pointing at object A and activating command X via the menu triggers a special response which may be different from the independent tasks of pointing at object A or activating command X.

Keyboard accelerators were introduced for use with desktop style menus so expert users could activate menu commands without moving their hands from the keyboard to the mouse. If the menu selection is known, a simple keystroke sequence can substitute for pulling down a menu and making a selection through the menu hierarchy. This type of interaction is impossible in the three-dimensional pop-up menu technique since there is no keyboard or suitable substitute. But using speech recognition, accelerators are automatically available. Any command can be activated at any time regardless of whether or not it is textually represented. As commands become more familiar, this technique is more commonly used.

The technique is flexible and allows for user preferences. For example, the menu bar can be placed at the top or the bottom of the display, font sizes and styles can be altered, and the menu can be removed if unwanted and recalled as required. The menu can be displayed as a default option or can be requested explicitly by the user if command prompting is necessary. For experienced users, the menu can be displayed on request and removed automatically once a command selection is made. This is not effective for novice users who need the menu more often since it would continually display and dismiss the menu unnecessarily. The most flexible method we have found is simply offering the user the ability to explicitly dismiss the menu and recall it when needed. A design principle we have followed is that the user should be allowed as much control as possible over the use of screen space.

Highlighting can be used as in conventional menus to show a selected item. Upon successful recognition of the verbal command, the item is visually highlighted briefly before either the next submenu is displayed or the command is performed.

In some cases, context sensitive menus or menu items are needed. This is easily achieved by exchanging the entire menu or disabling single items as they become unavailable. Since voice recognition is being used for input, a logical output mechanism is a synthetic speech system which could be used in much the same way that dialog boxes are used in traditional graphical user interfaces. For example, if a command requires that an object be selected for operation, voice feedback can prompt the user to make such a selection. This is most useful for textual output which cannot be displayed visually due to its length. However, it should be noted that lengthy verbal outputs are often inappropriate due to human memory abilities and voice intelligibility [9].

Due to the effective screen resolution discussed earlier, screen space is a scarce commodity and must be used judiciously. The menu must never dominate the display space. For this reason screen space will determine the breadth of the menu hierarchy. The menu can be confined to a single line or can be staggered over two lines as shown in figure 5. At an 18 point font size, it is possible to display the maximum recommended breadth of eight [10]. However, be-

MENU1 MENU3 MENU5
MENU2 MENU4

Figure 5: A two level, staggered menu configuration.

cause submenus do not show their relative position in the menu hierarchy, it is possible to become disoriented in the menu tree if it is deeper than three levels. We have specifically used a tree configuration (See figure 1.) to minimize this type of disorientation. It is important to remember to use the menu system sparingly in that not all tasks are best suited for this type of interaction.

The same metaphor of two-dimensional text over a three-dimensional world can be used for other purposes such as help information or a brief warning message. The text can be purposely centered and enlarged so as to attract the user’s attention if necessary or can be inconspicuous as is the case with the menu system. The recent use of two-dimensional text by Feiner et al. [3] shows more practical uses of text in three-dimensional worlds. This application can be easily extended from augmented reality to virtual worlds.

5. CONCLUSIONS

For conventional graphical user interfaces, pop-up and pull-down menus have been shown to be an effective technique for command selection. Utilizing human pointing abilities, this method provides clearly visible textual items with which selection tasks are trivial. These are desirable properties for menus in any environment. However, using a direct analogy from two-dimensional menus to three-dimensional worlds, we find that visibility is often poor and interaction is difficult. Visibility problems are primarily due to the fact that the menu does not always face the viewer. The user is free to move independently of the menu. Menu item selection is largely related to the performance of a six degree of freedom pointing task. Yet the selection is typically being made from a one or two-dimensional bank of items. Furthermore, the difficulty of this task is compounded by tracker noise and imprecision in the devices being utilized.

All too often, the use of the virtual worlds paradigm has implied the exclusive use of interaction techniques for three-dimensional spaces. We have argued here that there is a place in virtual worlds for two-dimensional interaction techniques. Menu interaction is one example. The hand is a primary interaction tool in virtual environments. Since textual interactions do not require the use of position or orientation information for which the hand is most useful, we have developed a technique using speech recognition to select items from a two-dimensional overlay on the three-dimensional world. This technique leaves the hands free for tasks demanding position and orientation specification. The menu does not move relative to the viewpoint. Text items always directly face the eyes. Menu item selection requires a simple spoken command which is presented on the menu itself. As users become more proficient with a particular menu hierarchy, the menu can be dismissed. The menu need not be present to enable item selection.

The hands-free interaction technique for menu selection illustrates the use of two-dimensional interaction techniques in three-dimensional spaces. Its use of a textual overlay on the viewplane has been shown to be an effective technique for textual interaction in virtual worlds. In the future, we expect to extend this metaphor by following our design philosophy of developing suitable interaction techniques which are dimensionally compatible with their corresponding tasks.

6. ACKNOWLEDGEMENTS

This research has been funded by the Effectiveness of Navy Electronic Warfare Systems (ENEWS) Program of the Tactical Electronic Warfare Division at the U.S. Naval Research Laboratory. The author wishes to thank Dr. Allen Duckworth and Dr. John Montgomery for their continued support of this work. Also, special thanks go to Dale Newcomb for his work with the speech recognition system and to Manuel Pérez and Tobin Hill for their invaluable assistance.

7. REFERENCES

[1] Callahan, J., Hopkins, D., Weiser, M., & Shneiderman, B. (1988). An Empirical Comparison of Pie vs. Linear Menus. *Proceedings of SIGCHI 1988*, 95—100.

[2] Company to Develop Speech Recognition Equipment for SGI. (1993, July). *Silicon Graphics World*, 3(7), p.11.

[3] Feiner, S., MacIntyre, B., Haupt, M., & Solomon, E. (1993). Windows on the World: 2D Windows for 3D Augmented Reality. *Proceedings of UIST '93,* 145—155.

[4] Foley, J.D., Wallace, V.L., & Chan, P. (1984). The Human Factors of Computer Graphics Interaction Techniques. *IEEE Computer Graphics and Applications,* 4(11), 13—48.

[5] Jacoby, R.H., & Ellis, S.R. (1992). Using Virtual Menus in a Virtual Environment. *Proceedings of the SPIE Technical Conference 1666.*

[6] Koved, L., & Shneiderman, B. (1986). Embedded Menus: Selecting Items in Context. *Communications of the ACM,* 29(4), 312—318.

[7] Nye, A. (1990). *Xlib Programming Manual for Version 11.* Sebastopol, CA.:O'Reilly & Associates.

[8] Perry, T.S., & Voelcker, J. (1989). Of Mice and Menus: Designing the User-Friendly Interface. *IEEE Spectrum*, September 1989, 46—51.

[9] Sanders, M.S., & McCormick, E.J. (1993). *Human Factors in Engineering and Design.* (7th ed.). New York: McGraw Hill, Inc.

[10] Shneiderman, B. (1987). *Designing the User Interface: Strategies for Effective Human-Computer Interaction.* Reading, MA.: Addison-Wesley.

[11] Smith, D.C., Irby, C., Kimball, R., Verplank, B., & Harslem, E. (1982). Designing the Star User Interface. *BYTE Magazine,* April 1982, 242—282.

[12] Somberg, B.L. (1987). A Comparison of Rule-Based and Positionally Constant Arrangements of Computer Menu Items. *Proceedings of SIGCHI 1987*, 255—260.

[13] Walker, N., & Smelcer, J.B. (1990). A Comparison of Selection Times from Walking and Pull-Down Menus. *Proceedings of SIGCHI 1990,* 221—225.